Combinatorial Mathematics

Mong-Jen Kao (高孟駿)

Monday 18:30 – 20:20

Outline

- Probabilistic Method – II
  - Linearity of Expectation
  - Large Deviation Inequalities
    - Markov’s Inequality, Chebyshev’s Inequality
    - The Chernoff Bounds
  - The Second Moment Method

Ex 1. Low-Degree Polynomials

The Prime Field $\mathbb{F}_2$

- Consider the prime field $\mathbb{F}_2 = \{0,1\}$.
  - We have the operators $+,\ -,\ \times,\ /$ defined over $\{0, 1\}$.
  - The result is to be mod by 2.

- For example,
  - $1 + 1 = 0$,
  - $0 + 1 = 1$,
  - $1 \times 0 = 0,\ \ 1 \times 1 = 1$, etc.

Polynomials over $\mathbb{F}_2$

- Consider the polynomial over $\mathbb{F}_2$.
  - A polynomial $f(x_1, \ldots, x_n)$ is said to have degree at most d if it has the following form

$$f(x_1, x_2, \ldots, x_n) \;=\; a_0 + \sum_{1 \le i \le m} \prod_{j \in S_i} x_j \;,$$

where $a_0 \in \{0,1\}$ and $S_i \subseteq [1, n]$ with $|S_i| \le d$.

Low-Degree Approximation for Products of Polynomials

- Intuitively, if $f_1, f_2, \ldots, f_m$ are polynomials of degree at most d, then $f \coloneqq \prod_{1 \le i \le m} f_i$ can have degree up to dm.

- The following lemma says that the product f can still be well-approximated by a low-degree polynomial.

Lemma 1 (Razborov 1987).

For any $r \geq 1$, there exists a polynomial g of degree at most dr such that $\Pr_{x \leftarrow \{0,1\}^n}[g(x) \neq f(x)] \leq 2^{-r}$.

Lemma 1 (Razborov 1987).

Let $f := \prod_{1 \le i \le m} f_i$,

where $f_1, f_2, \ldots, f_m$ are polynomials of degree at most d.

For any $r \ge 1$, there exists a polynomial g of degree at most dr such that

$$\Pr_{x \leftarrow \{0,1\}^n} [g(x) \neq f(x)] \le 2^{-r},$$

i.e., g and f differ on at most 2^{n-r} inputs.

- To prove Lemma 1, we define a random polynomial $g(x)$ and show that $\Pr[g(a) \neq f(a)] \le 2^{-r}$ holds for any input $a \in \{0,1\}^n$.

- To prove Lemma 1, we consider a random po… show that $\Pr[g(a) \neq f(a)] \leq 2^{-r}$ for any inpu…

Each possible subset is picked with probability 2^{-m}.

- Let $S_1, S_2, \ldots, S_r$ be random subsets sampled *independently* and *uniformly* from $\{1,2,\ldots,m\}$.

- Define

$$g \coloneqq \prod_{1 \leq j \leq r} h_j \,, \qquad \text{where } h_j \coloneqq 1 - \sum_{i \in S_j} (1 - f_i) \,.$$

- Let $S_1, S_2, \dots, S_r$ be random subsets sampled *independently* and *uniformly* from $\{1,2,\dots,m\}$.

- Define
$$g \coloneqq \prod_{1 \le j \le r} h_j\,, \qquad \text{where } h_j \coloneqq 1 - \sum_{i \in S_j} (1 - f_i)\;.$$

- Consider any input $a \in \{0,1\}^n$.
  - If $f(a) = 1$,
    then $f_i(a) = 1$ for all i, since $f = \prod_i f_i$.
    - Hence, $h_j(a) = 1$ for all j and
      $g(a) = 1 = f(a)$ with probability 1.

- Let $S_1, S_2, \ldots, S_r$ be random subsets sampled *independently* and *uniformly* from $\{1,2,\ldots,m\}$.
- Define $$g := \prod_{1 \le j \le r} h_j\,, \qquad \text{where } h_j := 1 - \sum_{i \in S_j} (1 - f_i)\ .$$
- Consider any input $a \in \{0,1\}^n$.
  - If $f(a) = 0$, then $f_i(a) = 0$ for at least one i.

    Let S' be the set of all such indexes.
    - By definition, $h_j(a) = 0$ if and only if S_j contains an odd number of elements from S'.

This happens with probability $1/2$.

- Consider any input $a \in \{0,1\}^n$.
  - If $f(a) = 0$, then $f_i(a) = 0$ for at least one i.

    Let S' be the set of all such indexes.

    This happens with probability $1/2$.

    - By definition, $h_j(a) = 0$ if and only if S_j contains an odd number of elements from S'.

    - Hence,

$$\Pr[\, g(a) = 0 \,] \;=\; 1 - \Pr\left[\, h_j(a) = 1 \; \forall j \,\right] \;=\; 1 - 2^{-r}\,.$$

- Consider any input $a \in \{0,1\}^n$.
  - If $f(a) = 1$, then $g(a) = f(a)$ for sure.
  - If $f(a) = 0$, then $g(a) = f(a)$ with probability $1 - 2^{-r}$.
- Let X_a be the indicator variable for the event that $g(a) \neq f(a)$ and $X \coloneqq \sum_a X_a$.
- We have $E[X] = \sum_a E[X_a] = \sum_a \Pr[X_a] \leq 2^{n-r}$.
  - Hence, there must exist such a collection of $S_1, \ldots, S_r$ such that $g(x)$ differs from $f(x)$ on at most 2^{n-r} inputs.

Large Deviation Inequalities

How Far can X Deviate from $\mathrm{E}[X]$?

- Expectation (expected value) is the *weighted average* of a variable taking a random value.
- Very often, knowing the expectation is not sufficient to know the true value of the variable.
  - Consider the random variable X that takes the values $\pm 10^{10}$ with probability $^1/_2$ each.
  - $E[X] = 0$, but X is either 10^{10} or -10^{10}.

Markov’s Inequality

- If $E[X]$ is what we only have,
  then a tight bound is given by the following theorem.

Theorem 2 (Markov’s Inequality).

If X is a non-negative random variable, then,

for any $t > 0$,

$$\Pr[\, X \geq t \,] \;\leq\; \frac{E[X]}{t} \;.$$

Alternatively, $\Pr[\, X \geq t \cdot E[X] \,] \;\leq 1/t$.

Theorem 2 (Markov's Inequality).

If X is a non-negative random variable, then,

for any $t > 0$,

$$\Pr[\, X \geq t \,] \;\leq\; \frac{E[X]}{t} \,.$$

- We have

$$E[X] \;=\; \sum_{i} i \cdot \Pr[X = i] \;\geq\; \sum_{i \geq t} t \cdot \Pr[X = i] \;=\; t \cdot \Pr[X \geq t] \;.$$

- The above bound is tight,

  if $E[X]$ is what we only have.

Chebyshev's Inequality

- If we also know $\mathrm{Var}[X]$,
  then a (much) tighter guarantee can be obtained.

<u>**Theorem 2 (Chebyshev's Inequality).**</u>

For any $t > 0$,

$$\Pr[\,|X - E[X]| \geq t\,] \;\leq\; \frac{\mathrm{Var}[X]}{t^2}\;.$$

Alternatively,

$$\Pr\left[\,|X - E[X]| \;\geq\; t \cdot \sqrt{\mathrm{Var}[X]}\,\right] \leq\; 1/t^2\;.$$

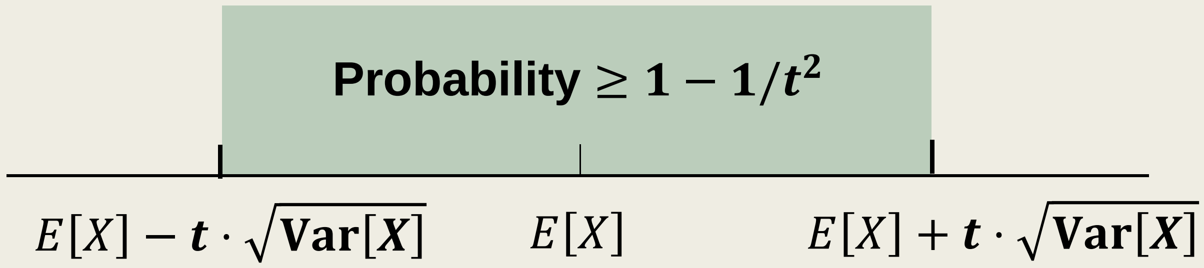
Theorem 2 (Chebyshev's Inequality).

For any $t > 0$,

$$\Pr[\, |X - E[X]| \geq t \,] \;\leq\; \frac{\mathrm{Var}[X]}{t^2} \;.$$

- Consider the random variable $Y \coloneqq (X - E[X])^2 \geq 0$.
  - Apply the Markov's inequality, we obtain

$$\Pr[\, |X - E[X]| \geq t \,] \;=\; \Pr[\, Y \geq t^2 \,] \;\leq\; \frac{E[\,(X - E[X])^2\,]}{t^2} \;=\; \frac{\mathrm{Var}[X]}{t^2} \;.$$



Moment Generating Function

&

The Chernoff Bounds

Moments of a Random Variable

- The k^{th} moment of a random variable X is defined as $E\left[X^k\right]$.
  - The 1^{st}-moment is exactly the expectation $E[X]$.
  - The 2^{nd}-moment gives the variance

$$\mathrm{Var}[X] \;\coloneqq\; E[(X - E[X])^2] \;=\; E[X^2] - (E[X])^2 \;.$$

The Moment Generating Function

- The moment generating function of a random variable X is defined as
$$M_X(t) \coloneqq E[\, e^{tX} \,]\,.$$

- The moment generating function $M_X(t)$ is important in that
  - It captures all the moments of X.
  - We have
$$E[X^n] = M_X^{(n)}(0)\,,$$
    where $M_X^{(n)}(t)$ is the n^{th}-derivative of $M_X(t)$.

The Chernoff Bounds

- If we have the mgf $M_X(t)$ of X, then the tightest *concentration bound* is given by the Chernoff bounds.

Theorem 3 (Chernoff Bounds).

For any $t > 0$,

$$\Pr[\, X \geq a \,] \;=\; \Pr[\, e^{tX} \geq e^{ta}] \;\leq\; E[e^{tX}] \cdot e^{-ta}.$$

Similarly, for any $t < 0$,

$$\Pr[\, X \leq a \,] \;=\; \Pr[\, e^{tX} \geq e^{ta}] \;\leq\; E[e^{tX}] \cdot e^{-ta}.$$

The Chernoff Bounds

- If we have the mgf $M_X(t)$ of X, then the tightest *concentration bound* is given by the Chernoff bounds.

- Theorem 3 gives the original form of Chernoff bounds, which is derived from the Markov's inequality.
  - Depending on what the actual distribution of X, the Chernoff bounds may have different final form.

  - As an example, let's consider the *sum of independent variables from* $[0,1]$.

Theorem 4 (Chernoff Bounds for Sum of Independent Variables).

Let $X_1, X_2, \dots, X_n$ be independent variables taking values from the interval $[0,1]$. Let $X \coloneqq \sum_i X_i$ and $\mu \coloneqq E[X]$.

Then, for any $a > 0$,

$$\Pr[\, X \geq \mu + a \,] \leq \; e^{-\frac{a^2}{2n}} \quad \text{and} \quad \Pr[\, X \geq \mu - a \,] \leq \; e^{-\frac{a^2}{2n}} \; .$$

- Intuitively, the bound says that the outcome of X concentrates between $\mu \pm \theta(\sqrt{n})$.
  - Outside this interval, the likelihood decreases *exponentially*.

Theorem 4 (Chernoff Bounds for Sum of Independent Variables).

Let $X_1, X_2, \dots, X_n$ be independent variables taking values from the interval $[0,1]$. Let $X \coloneqq \sum_i X_i$ and $\mu \coloneqq E[X]$.

Then, for any $a > 0$,

$$\Pr[\, X \geq \mu + a \,] \leq e^{-\frac{a^2}{2n}} \quad \text{and} \quad \Pr[\, X \geq \mu - a \,] \leq e^{-\frac{a^2}{2n}} .$$

- Taking $t = O\left(\sqrt{n \ln n}\right)$,

  the above probability is bounded by $O(n^{-1})$.

The Second Moment Method

The Second Moment Method

- Let X be a non-negative integer-valued random variable.
- The following inequality, obtained from Chebyshev's inequality, is one typical way and often useful.

$$\Pr[\, X = 0 \,] \;\le\; \frac{\mathrm{Var}[X]}{(E[X])^2}\,.$$

  – Indeed, we have

$$\Pr[X = 0] \;\le\; \Pr[\, |X - E[X]| \ge E[X] \,] \;\le\; \mathrm{Var}[X] \,/\, (E[X])^2\,.$$

Ex 2. Threshold Behavior in Random Graphs

The Random Graph $G_{n,p}$

- Consider the graph $G_{n,p} = (V, E)$ with $|V| = n$ and the edge set E generated randomly as follows.
  - For any $u, v \in V$, we draw an edge $(u, v) \in E$ *independently* with probability p.

- It follows that

$$\mathrm{E}[\,|E|\,] = \binom{n}{2}p \quad \text{and} \quad \Pr[\,|E| = m\,] \;=\; p^m(1-p)^{\binom{n}{2}-m}\,.$$

The Threshold Behavior of $G_{n,p} \supseteq K_4$

- Let G be a realization (sample) of $G_{n,p}$ and consider the event that G contains a clique of size 4.

- We have the following theorem.

Theorem 5. For any $\epsilon > 0$ and *sufficiently large* n,

if $p = o\left(n^{-2/3}\right)$, then

$$\Pr[\ G \text{ contains } K_4\ ] < \epsilon\ .$$

On the contrary, if $p = \omega\left(n^{-2/3}\right)$, then

$$\Pr[\ G \text{ does not contain } K_4\ ] < \epsilon\ .$$

Theorem 5. For any $\epsilon > 0$ and *sufficiently large* n,

if $p = o\left(n^{-2/3}\right)$, then $$\Pr[\ G \text{ contains } K_4\ ] < \epsilon\ .$$

- Suppose that $p = o\left(n^{-2/3}\right)$.
  - Let $C_1, \ldots, C_{\binom{n}{4}} \subseteq V$ be all possible set of 4 vertices in G.
  - Let $X_i = \begin{cases} 1 & \text{if } C_i \text{ is a } K_4, \\ 0 & \text{otherwise,} \end{cases}$ and $X \coloneqq \sum_i X_i$.
- It follows that $\Pr[X_i] = p^6 = o(n^{-4})$ and $E[X] = \binom{n}{4} o(n^{-4}) = o(1)$.
- Since X is integer-valued, $\Pr[X \geq 1] \leq E[X] < \epsilon$ for sufficiently large n.

Theorem 5. For any $\epsilon > 0$ and *sufficiently large* n, if $p = \omega\left(n^{-2/3}\right)$, then

$$\Pr[\ G \text{ does not contain } K_4\ ] < \epsilon\ .$$

- Suppose that $p = \omega\left(n^{-2/3}\right)$.
  - In this case $E[X] \to \infty$ as n tends to infinity.
  - This, however, is *not strong enough* to guarantee the statement of the theorem.
- We will show that $\mathrm{Var}[X] = o\left((E[X])^2\right)$.
  - Then we have $\Pr[X = 0] = o(1)$ and the theorem holds.

- Suppose that $p = \omega\left(n^{-2/3}\right)$.
  - We will show that $\mathrm{Var}[X] = o\left((E[X])^2\right)$.

- To compute $\mathrm{Var}[X]$, we need the following lemma.

<u>Lemma 6.</u>

Let $Y_1, \ldots, Y_m$ be 0-1 random variable and $Y \coloneqq \sum_i Y_i$.

Then $$\mathrm{Var}[\,Y\,] \le E[Y] + \sum_{\substack{1 \le i,j \le m,\\ i \ne j}} \mathrm{Cov}\left(Y_i, Y_j\right),$$

where $\mathrm{Cov}\left(Y_i, Y_j\right) \coloneqq E\left[\,Y_i \cdot Y_j\,\right] - E[Y_i] \cdot E\left[Y_j\right]$.

- Suppose that $p = \omega\left(n^{-2/3}\right)$.
  - We will show that $\mathrm{Var}[X] = o\left((E[X])^2\right)$.

- For any $1 \le i, j \le m$ with $i \ne j$,

  consider the covariance of X_i and X_j.

  - If $\left|C_i \cap C_j\right| \le 1$,

    then C_i and C_j share no edge, and X_i and X_j are independent.

    Hence, $E\left[X_i X_j\right] = E[X_i] \cdot E\left[X_j\right]$ and $\mathrm{Cov}\left(X_i, X_j\right) = 0$.

- For any $1 \leq i, j \leq m$ with $i \neq j$, consider the covariance of X_i and X_j.

  – If $|C_i \cap C_j| = 2$, then C_i and C_j ***share one edge***.

    The *11 edges* in $C_i \cup C_j$ have to appear at the same time for $X_i \cdot X_j$ to be 1.

    Hence,

    $$\mathrm{Cov}(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j] \leq E[X_i X_j] = p^{11}.$$

    There are $\binom{n}{6} \cdot \binom{6}{2;2;2}$ such pairs of C_i and C_j.

- For any $1 \le i, j \le m$ with $i \ne j$,
  consider the covariance of X_i and X_j.

  – Similarly, if $|C_i \cap C_j| = 3$, then C_i and C_j ***share three edges***.

    The *9 edges* in $C_i \cup C_j$ have to appear at the same time for $X_i \cdot X_j$ to be 1.

    Hence,

$$\mathrm{Cov}(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j] \le E[X_i X_j] = p^9 .$$

    There are $\binom{n}{5} \cdot \binom{5}{1;3;1}$ such pairs of C_i and C_j.

- For any $1 \leq i, j \leq m$ with $i \neq j$, consider the covariance of X_i and X_j.
  - Apply Lemma 6, we obtain

$$\mathrm{Var}[X] \;\leq\; E[X] + \sum_{i \neq j} \mathrm{Cov}\left(X_i, X_j\right)$$

$$\leq \binom{n}{4} p^6 + \binom{n}{6} \cdot \binom{6}{2;2;2} p^{11} + \binom{n}{5} \cdot \binom{5}{1;3;1} \mathrm{p}^9$$

$$= \theta\left(n^6 p^{11}\right)$$

$$= o\left((E[X])^2\right) \;\text{ since } (E[X])^2 = \theta\left(n^8 p^{12}\right) \text{ and } p = \omega\left(n^{-2/3}\right).$$

- It remains to prove the following lemma.

> **Lemma 6.**
>
> Let $Y_1, \dots, Y_m$ be 0-1 random variable and $Y \coloneqq \sum_i Y_i$.
>
> Then $$\mathrm{Var}[\,Y\,] \;\leq\; E[Y] + \sum_{\substack{1\leq i,j\leq m,\\ i\neq j}} \mathrm{Cov}(Y_i, Y_j)\ .$$

- By definition, we have $\mathrm{Var}[Y] = \sum_i \mathrm{Var}[Y_i] + \sum_{i\neq j} \mathrm{Cov}(Y_i, Y_j)$.
  - Since Y_i is a 0-1 random variable, $E\left[Y_i^2\right] = E[Y_i]$.
  - Hence, $\mathrm{Var}[Y_i] = E\left[Y_i^2\right] - (E[Y_i])^2 \;\leq\; E[Y_i]$.